

--	--	--	--	--	--	--	--	--

# MULTIMEDIA UNIVERSITY

## FINAL EXAMINATION

TRIMESTER 1, 2018/2019

### HPB3011 –BIOINFORMATICS ALGORITHMS II

(All sections / Groups)

19 October 2018

3:00 -5:00 PM

(2 hours)

---

#### INSTRUCTIONS TO STUDENTS

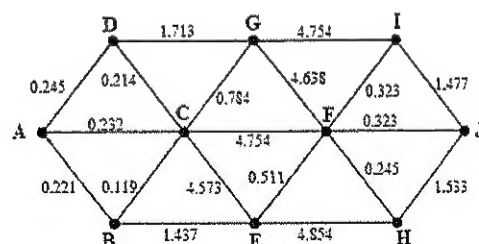
1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use a calculator.

**Question 1**

- a) Why is it necessary to normalize DNA microarray data? [2 marks]
- b) What are the two main types of microarray gene chips? [1 mark]
- c) In the whole transcriptome analysis of eukaryotes, RNA-seq read mapper such as Tophat or HiSat is more favorable over other mapping algorithms as implemented in Burrow-wheeler Aligner and Bowtie. What is the possible reason contributing to the tool selections in the analysis? [1 mark]
- d) Why is RNA study still important despite extensive research on genomics? [1 mark]
- e) In the effort of understanding the expression of *Zingiber officinale* (ginger), the transcriptome was sequenced using single-end Illumina technology. However, evaluation on the mapping rate is not satisfactory. What could be the possible reason? [2 marks]
- f) Give TWO advantages of RNA-seq compared to DNA microarray for the study of transcriptomics. [2 marks]
- g) One of the challenges in RNA High Throughput Sequencing (HTS) is the generation of fragmented transcriptome through RNA assembly. Suggest one method to capture transcriptome and retaining the sequence completeness. [1 mark]

**Question 2**

- a) A diagnostic test is 92% sensitive and 84% specific. A test group is comprised of 500 people known to have the disease and 500 people known to be free of the disease. How many of the known positives would actually test positive? How many of the known negatives would actually test negative? [2 marks]
- b) The graph below illustrate the Euclidean distances among the genes (A,B,C,D,E,F,G,H,I,J). Based on the information given, what is the appropriate clustering method should be used? Draw the final graph and state the total distance of the tree. [3 marks]



Start = any point

Continued .....

- c) Given the classification in Figure I below, what decide the position of the hyperplane for the separation of the classes? What will happen to the hyperplane when a data is added, as illustrated in Figure II. [2 marks]

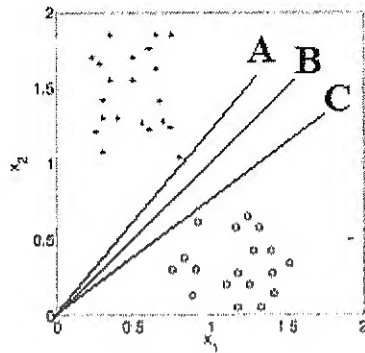


Figure 1

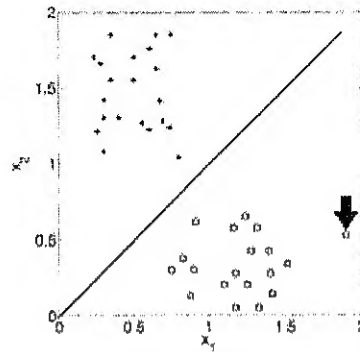


Figure 2

- d) Explain the 3 main approaches in markov model. [3 marks]

### Question 3

- a) What is the concept of Chou-Fasman in protein secondary structure prediction? How is Chou-Fasman algorithms different than those in GOR methods? [2 marks]
- b) What are the two artificial neural levels in PHD for secondary structure prediction? Explain the differences. [4 marks]
- c) A researcher intend to study the functional orthology in a newly isolated protein. Suggest any two databases for the purpose. Why databases in NCBI and Uniprot may not be suitable for the analysis? [2 marks]
- d) Describe Levinthal paradox. [2 marks]

Continued .....

**Question 4**

- a) Given the matrix below, draw the structure of the RNA based on Nussinov algorithm. [2 marks]

	C	C	C	U	U	U	U	A	G	G
C	0	0	0	0	0	0	0	1	2	3
C		0	0	0	0	0	0	1	2	3
C			0	0	0	0	0	1	2	2
U				0	0	0	0	1	1	1
U					0	0	0	1	1	1
U						0	0	1	1	1
U							0	1	1	1
A								0	0	0
G									0	0
G										0

- b) Compare Mutual Information and Co-variance Scoring in RNA structure prediction. [3 marks]
- c) One of the functions of pseudoknot is to serve as chaperone. How is the chaperone function different than ubiquitin? [1 mark]
- d) List any 2 classes of RNA and their functions. [1 mark]
- e) What are the three limitations of Nussinov Algorithm? [3 marks]

**END OF PAPER**